

Class 11: AlphaFold

AUTHOR

Sarah Mirsaidi Madjdabadi, A16890186

Here we read the results from AlphaFold and try to interpret all the models and quality score metrics:

```
library(bio3d)

pth <- "dimer_23119/"
pdb.files <- list.files(path = pth, full.names = TRUE, pattern
```

Align and superpose all these models

```
file.exists(pdb.files)
```

```
[1] TRUE TRUE TRUE TRUE TRUE
```

```
pdbs <- pdbaln(pdb.files, fit = TRUE, exefile="msa")
```

Reading PDB files:

```
dimer_23119//dimer_23119_unrelaxed_rank_001_alphafold2_multime
r_v3_model_2_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_002_alphafold2_multime
r_v3_model_5_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_003_alphafold2_multime
r_v3_model_4_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_004_alphafold2_multime
r_v3_model_1_seed_000.pdb
dimer_23119//dimer_23119_unrelaxed_rank_005_alphafold2_multime
r_v3_model_3_seed_000.pdb
.....
```

Extracting sequences

```
pdb/seq: 1   name:
dimer_23119//dimer_23119_unrelaxed_rank_001_alphafold2_multime
r_v3_model_2_seed_000.pdb
pdb/seq: 2   name:
dimer_23119//dimer_23119_unrelaxed_rank_002_alphafold2_multime
r_v3_model_5_seed_000.pdb
```

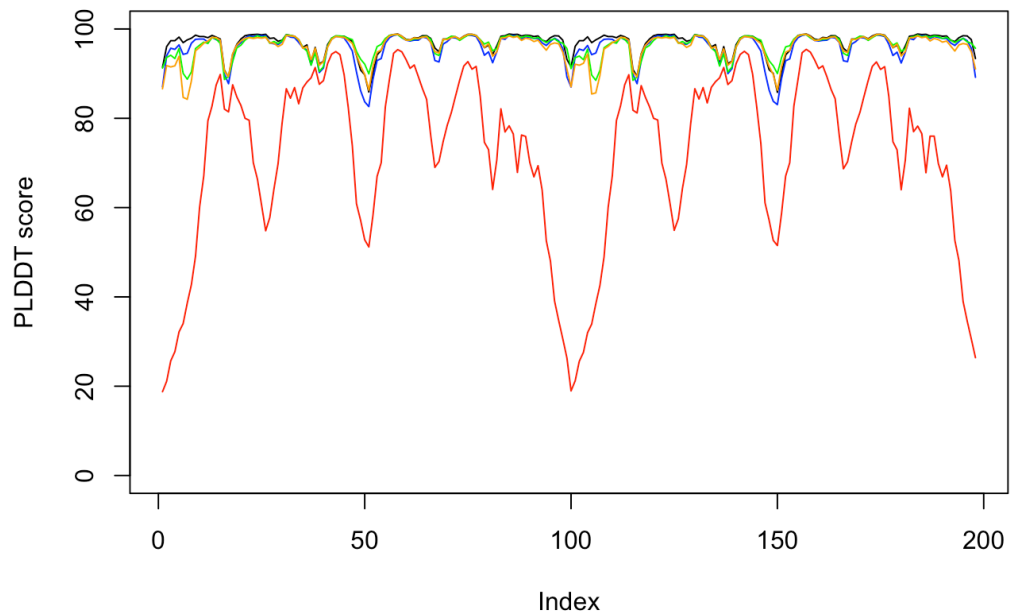
```
pdb/seq: 3   name:  
dimer_23119//dimer_23119_unrelaxed_rank_003_alphafold2_multime  
r_v3_model_4_seed_000.pdb  
pdb/seq: 4   name:  
dimer_23119//dimer_23119_unrelaxed_rank_004_alphafold2_multime  
r_v3_model_1_seed_000.pdb  
pdb/seq: 5   name:  
dimer_23119//dimer_23119_unrelaxed_rank_005_alphafold2_multime  
r_v3_model_3_seed_000.pdb
```

```
library(bio3dview)
```

```
view.pdbs(pdbs)
```



```
plot(pdbs$b[1,], typ="l", ylim=c(0,100), ylab="PLDDT score")  
lines(pdbs$b[2,],typ="l", col="blue")  
lines(pdbs$b[3,],typ="l", col="green")  
lines(pdbs$b[4,],typ="l", col="orange")  
lines(pdbs$b[5,],typ="l", col="red")
```

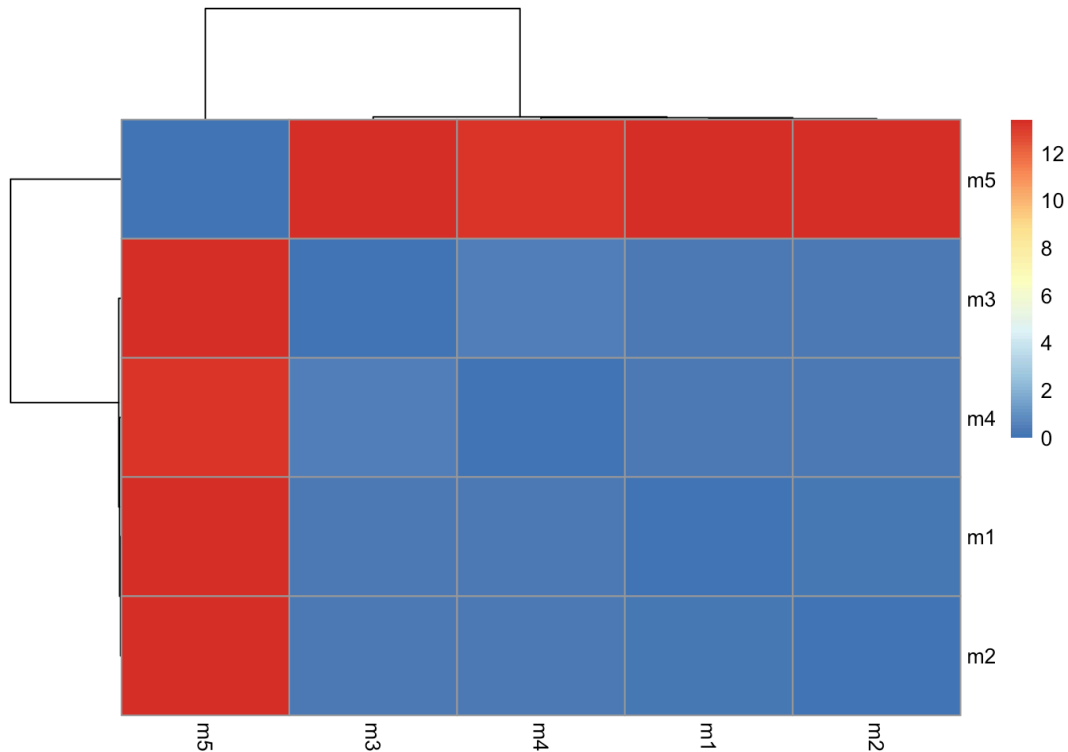


```
rd <- rmsd(pdb)
```

Warning in rmsd(pdb): No indices provided, using the 198 non NA positions

```
library(pheatmap)

colnames(rd) <- paste0("m",1:5)
rownames(rd) <- paste0("m",1:5)
pheatmap(rd)
```



```
library(jsonlite)
```

```
# Listing of all PAE JSON files
pae_files <- list.files(path=pth,
                        pattern=".*model.*\\.json",
                        full.names = TRUE)
```

```
pae1 <- read_json(pae_files[1],simplifyVector = TRUE)
pae2 <- read_json(pae_files[2],simplifyVector = TRUE)
pae3 <- read_json(pae_files[3],simplifyVector = TRUE)
pae4 <- read_json(pae_files[4],simplifyVector = TRUE)
pae5 <- read_json(pae_files[5],simplifyVector = TRUE)
```

```
attributes(pae1)
```

```
$names
```

```
[1] "plddt" "max_pae" "pae" "ptm" "iptm"
```

```
# Per-residue pLDDT scores
# same as B-factor of PDB..
head(pae1$plddt)
```

```
[1] 91.44 96.06 97.38 97.38 98.19 96.94
```

```
pae1$max_pae
```

```
[1] 13.57812
```

```
pae2$max_pae
```

```
[1] 15.71094
```

```
pae3$max_pae
```

```
[1] 12.41406
```

```
# Best (lowest) PAE score
```

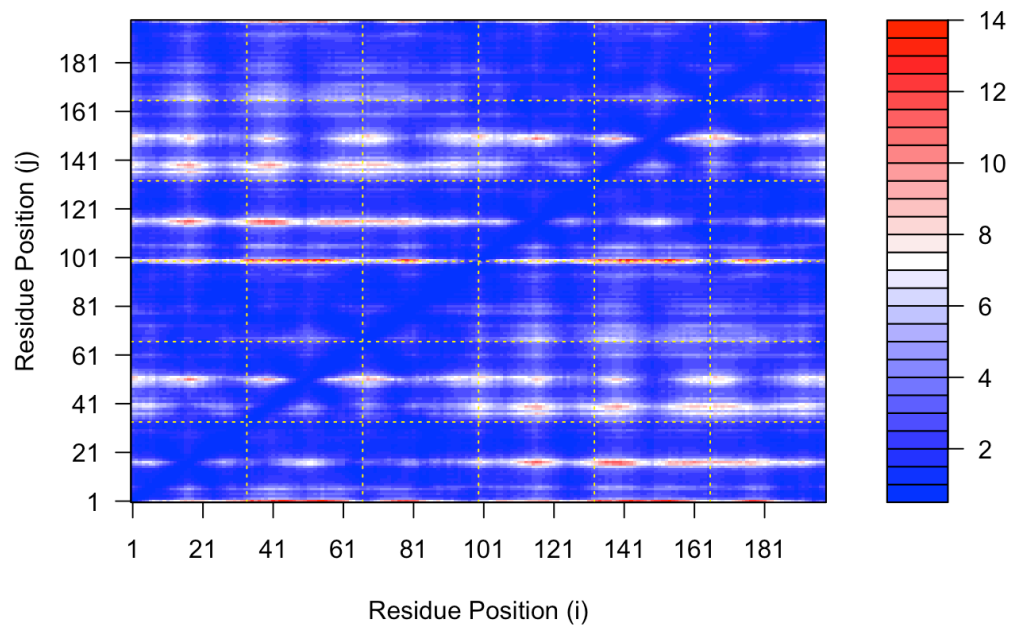
```
pae4$max_pae
```

```
[1] 19.95312
```

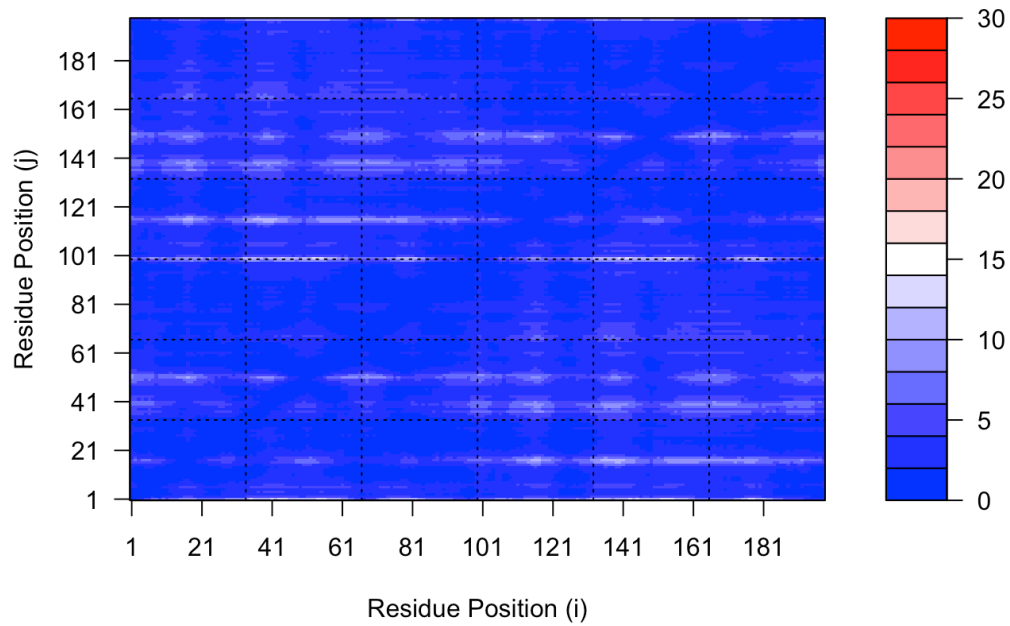
```
pae5$max_pae
```

```
[1] 29.85938
```

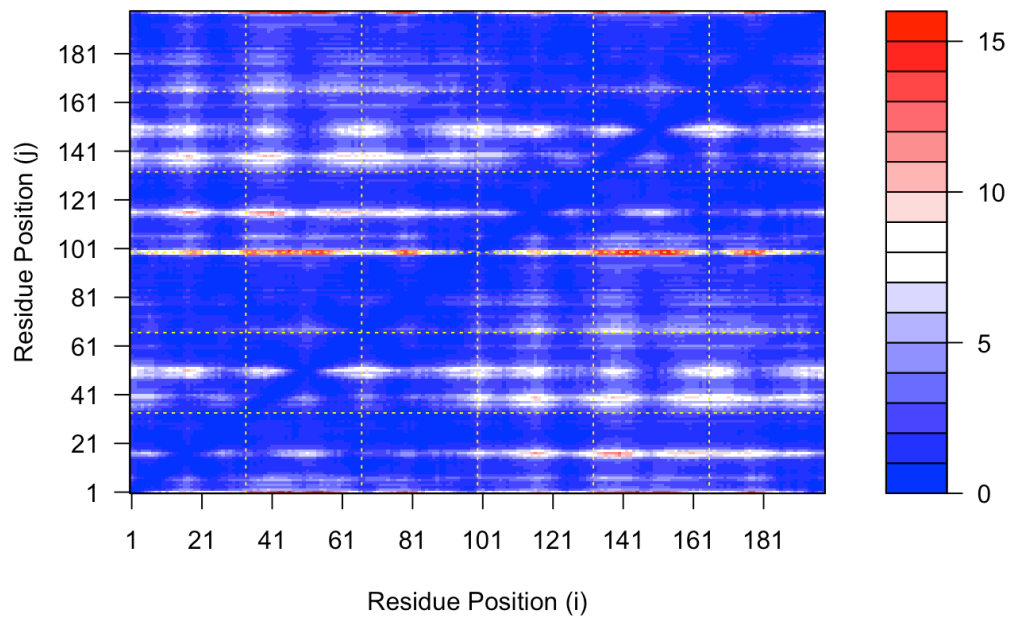
```
plot.dmat(pae1$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)")
```



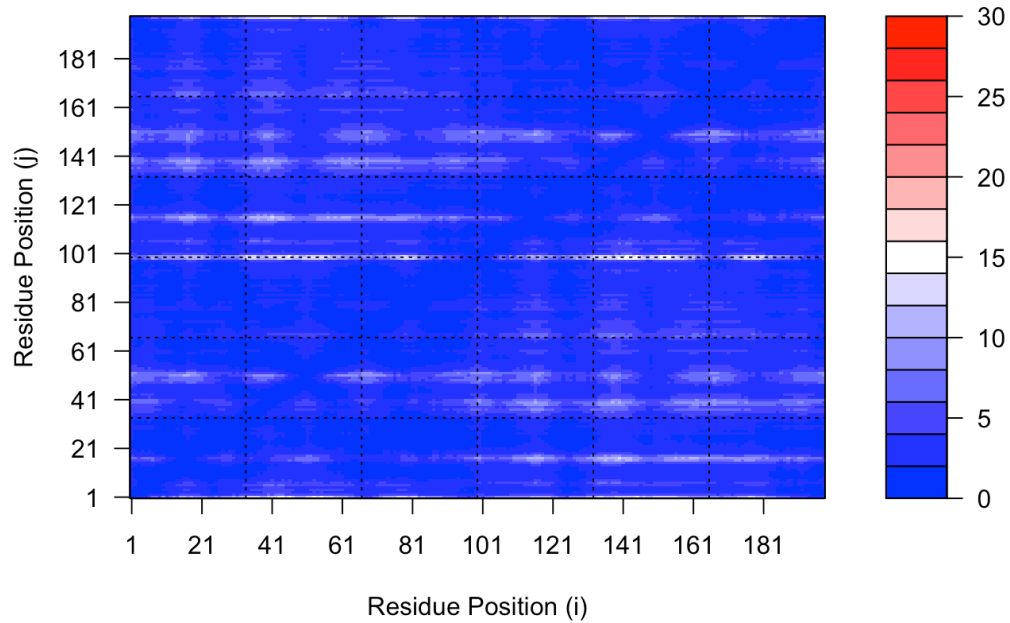
```
plot.dmat(pae1$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)",  
          grid.col = "black",  
          zlim=c(0,30))
```



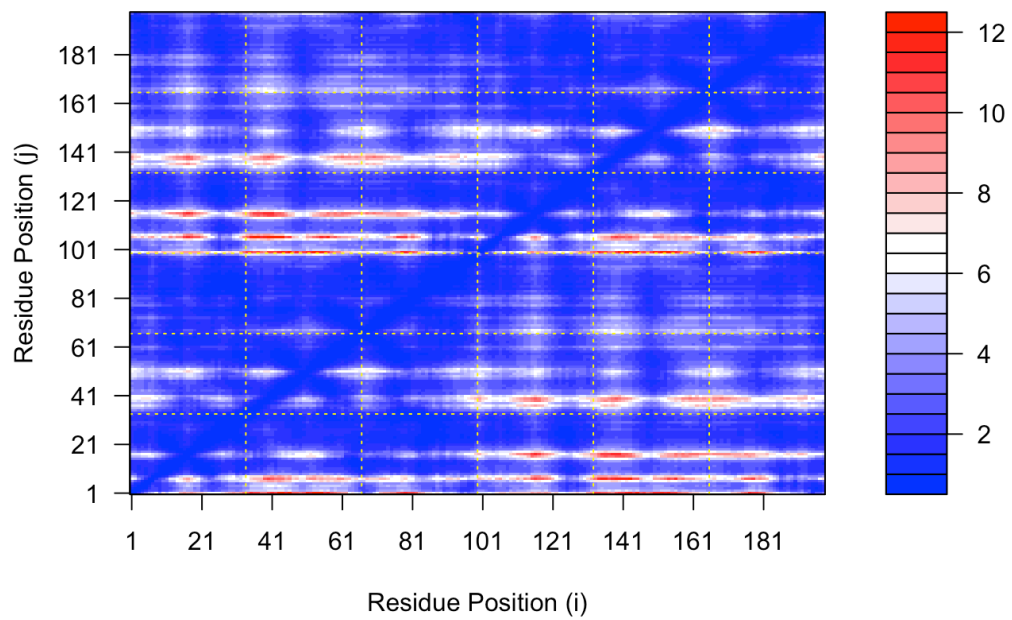
```
plot.dmat(pae2$paes,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)")
```



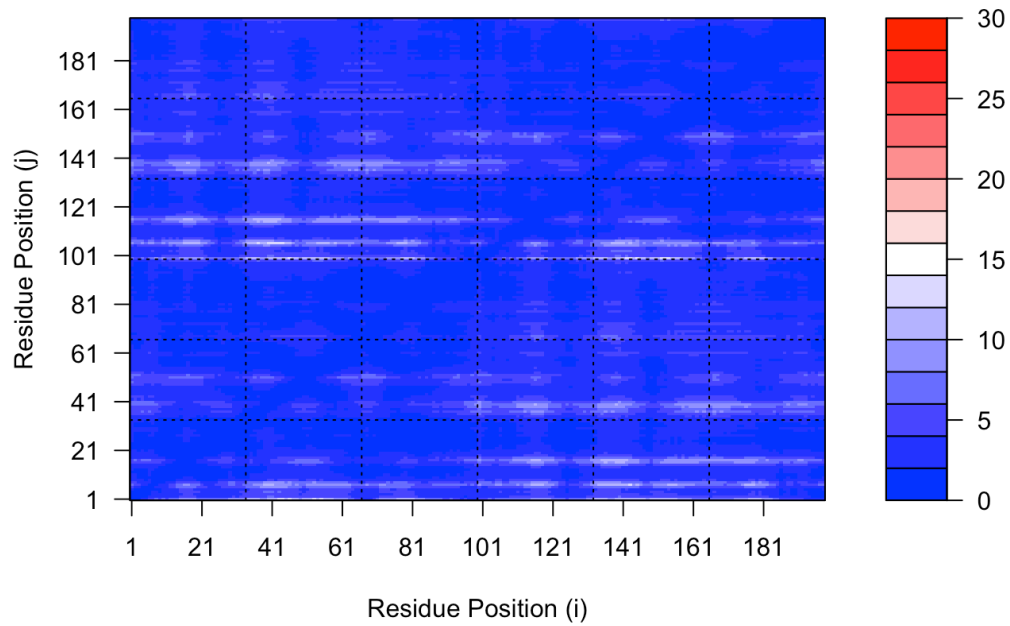
```
plot.dmat(pae2$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)",  
          grid.col = "black",  
          zlim=c(0,30))
```



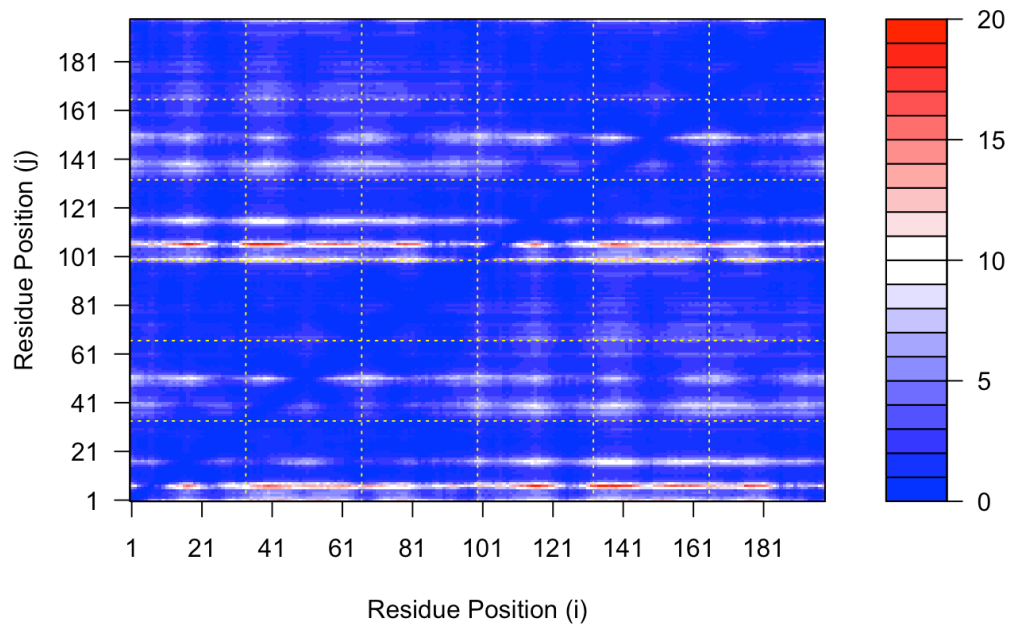
```
plot.dmat(pae3$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)")
```

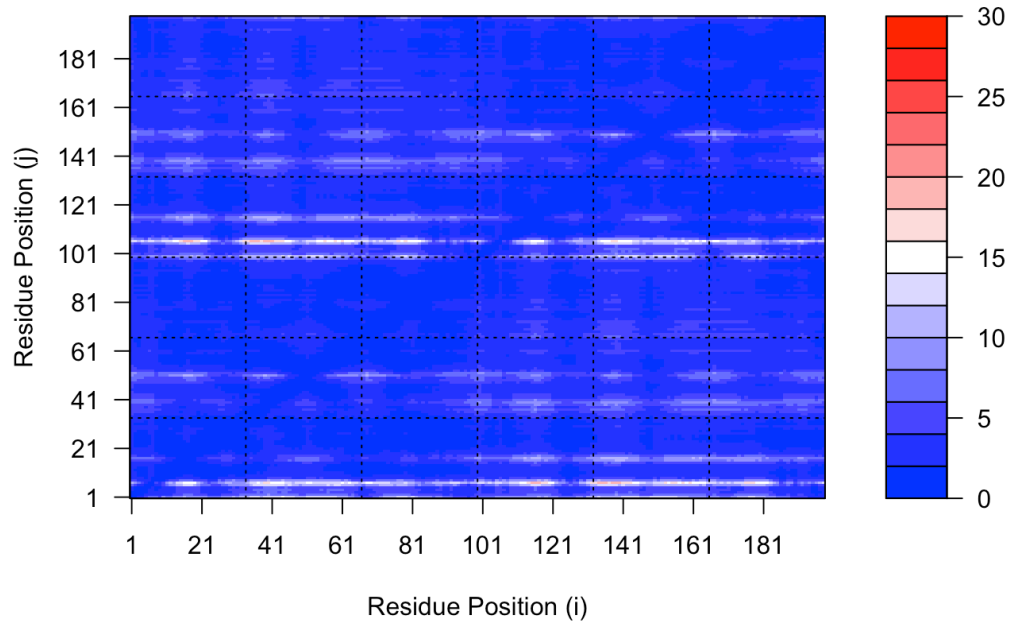
```
plot.dmat(pae3$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)",  
          grid.col = "black",  
          zlim=c(0,30))
```



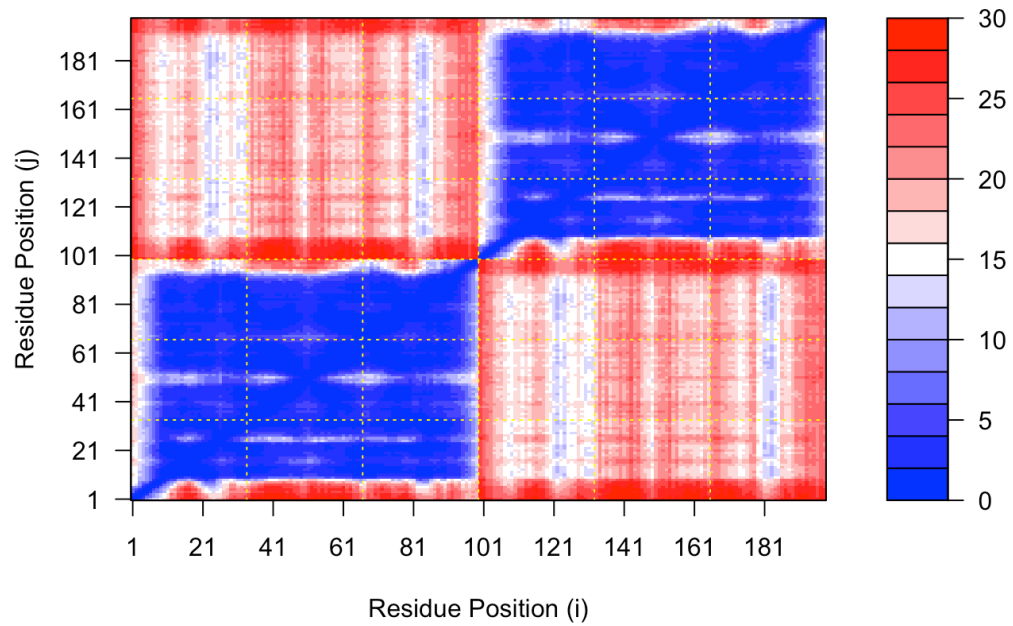
```
plot.dmat(pae4$paes,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)")
```



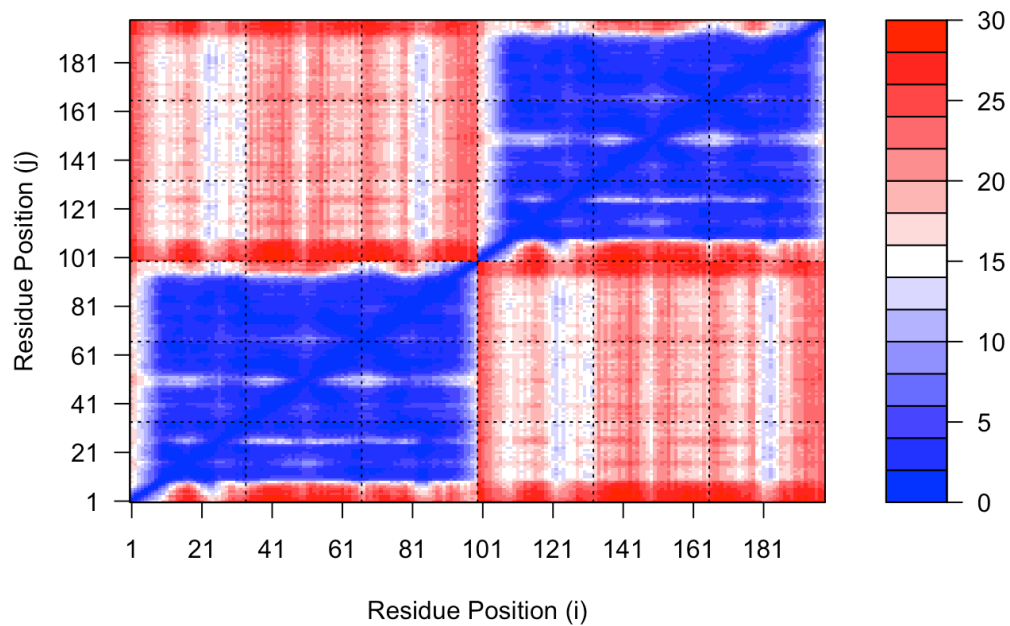
```
plot.dmat(pae4$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)",  
          grid.col = "black",  
          zlim=c(0,30))
```



```
plot.dmat(pae5$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)")
```



```
plot.dmat(pae5$pae,  
          xlab="Residue Position (i)",  
          ylab="Residue Position (j)",  
          grid.col = "black",  
          zlim=c(0,30))
```



Residue conservation from alignment file

```
aln_file <- list.files(path=pth,
                      pattern=".a3m$",
                      full.names = TRUE)

aln_file
```

```
[1] "dimer_23119//dimer_23119.a3m"
```

```
aln <- read.fasta(aln_file[1], to.upper = TRUE)
```

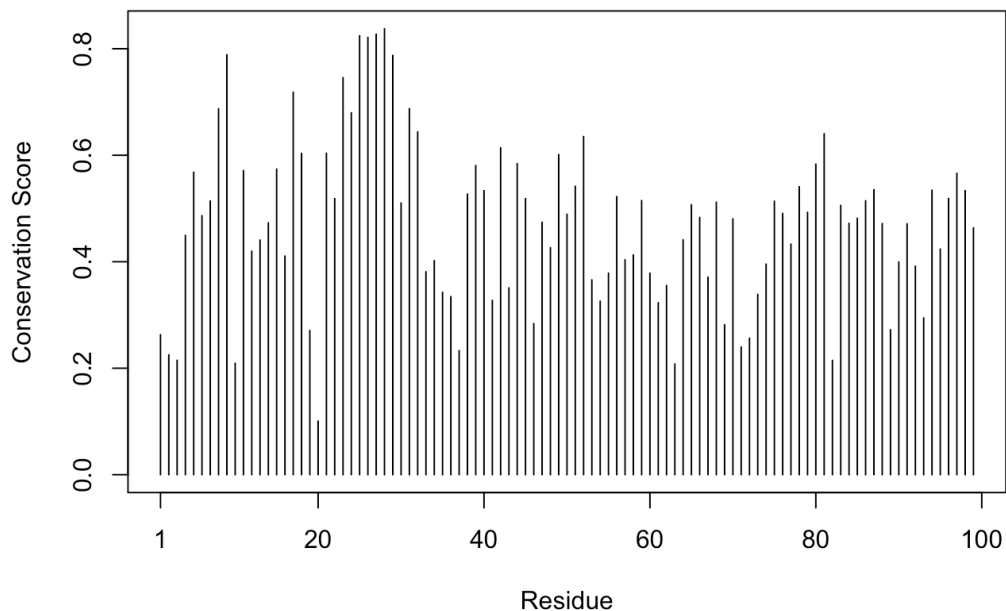
```
[1] " ** Duplicated sequence id's: 101 **"
[2] " ** Duplicated sequence id's: 101 **"
```

```
dim(aln$ali)
```

```
[1] 5378 132
```

```
sim <- conserv(aln)
```

```
plotb3(sim[1:99],
       ylab="Conservation Score")
```



```
con <- consensus(aln, cutoff = 0.9)
con$seq
```

```
[1] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
    "-" "-" "-" "-"
[19] "-" "-" "-" "-" "-" "-" "D" "T" "G" "A" "-" "-" "-" "-"
    "-" "-" "-" "-"
[37] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
    "-" "-" "-" "-"
[55] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
    "-" "-" "-" "-"
[73] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
    "-" "-" "-" "-"
[91] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
    "-" "-" "-" "-"
[109] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"
    "-" "-" "-" "-"
[127] "-" "-" "-" "-" "-" "-"
```

```
m1.pdb <- read.pdb(pdb.files[1])  
occ <- vec2resno(c(sim[1:99], sim[1:99]), m1.pdb$atom$resno)  
write.pdb(m1.pdb, o=occ, file="m1_conserv.pdb")
```